International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 6, Nov-Dec-2024, ISSN (Online): 2395-566X

Full Stack Web Application for Prediction and Diagnosis of Heart Disease

Assistant Professor Ms. Dornadhula Danya, Suraj A U,
Moju Kumar B L, Deepak Kumar Singh D, Shubhan GC
School of Information Science Presidency University Bangalore, India.

Abstract- In the modern era, Cardio-vascular disease has high prevalence and rate of mortality which proves how critical, identification and intervention strategies are, further highlighting the importance of incorporating this in developing heart disease prediction systems. The heart prediction system research revolves around using AI-driven techniques techniques to strengthen and make heart disease risk prediction robust and effective. The paper explains methodology, dataset characteristics, experimental setup, results and the design of the models in a AI-driven techniques heart prediction system. Additionally, the practical implications of the research output are discussed regarding the use of the system in real life for alleviating heart disease predictions and strategies.

Index Terms- Prediction, Heart Disease, Web Application

I. INTRODUCTION

Cardiovascular disease continues to be a leading cause of death worldwide, highlighting the need for early detection and timely medical intervention. The condition encompasses various disorders, such as coronary artery disease (CAD), and can manifest through symptoms such as chest pain, shortness of breath, and irregular heartbeats. If left undiagnosed, these symptoms can progress to severe complications like heart attacks or heart failure.

The challenge in diagnosing heart disease lies in the overlap of symptoms and the presence of risk factors such as age, gender, blood pressure, cholesterol levels, smoking habits, history. While traditional diagnostic methods remain effective, they often lack the speed, scalability, or precision needed for large- scale screening of the population.

In this research, we present a system that utilizes the a robust ensemble algorithm classifier, a powerful AI-driven techniques algorithm known for its high accuracy and ability to handle complex, high-dimensional data. Our model is trained on clinical datasets containing key features related to cardiovascular health, allowing for automated and precise predictions of heart disease. The goal is to improve the early detection of heart disease, leading to faster intervention and better patient outcomes. Our study emphasizes the role of AI-driven techniques in healthcare, particularly for predictive diagnostics. The results demonstrate the effectiveness of the a robust ensemble algorithm model in classifying heart disease risk factors, providing a scalable and reliable tool for heart disease prediction. This research contributes to the development of efficient, accessible diagnostic tools for healthcare providers, helping to improve the identification and treatment of cardiovascular conditions.

II. PROBLEM STATEMENT

Detection and Monitoring a Cardiovascular disease is one of the most challenging tasks in the healthcare system. Although there are some coronary disease prediction tools, however, some are relatively very costly and others do not even have minimum accuracy for normal use. As per studies and available data, it is understood that identifying heart diseases at an early stage is crucial in decreasing the number of deaths and managing the symptoms well. However, the constant observation and regular check-ups of the patient in many instances are not feasible because it requires a lot of time, knowledge and resources. In an era where data is so accessible, utilizing AI-driven techniques methods does seem helpful as one can analyze a lot of data and look for useful patterns. These patterns can be used as an important source of identification in the field of medicine. Such technologies, especially those based on Python programming can provide effective solutions in computational analysis of medical data and improve efficiency for the field of medicine.

As one of the programming languages that are widely applied in the healthcare analytics sector, Python has good libraries and tools for the analysis of medical datasets. High levels of so-called "bad" LDL cholesterol is one of the leading causes of coronary artery disease being the most widespread heart related disease. The disease is often accompanied with symptoms including chest pain, shortness of breath and so on.

III. LITERATURE REVIEW

Cardiovascular disease prediction has been a focal point of medical research, leveraging advancements in AI-driven techniques to improve diagnostic accuracy and reliability.

Techniques Used
Ensemble Models
Robust ensemble algorithm, as implemented in your project, is a leading algorithm in this domain due to its ability to handle diverse datasets and effectively capture feature interactions. It is particularly valued for its robustness, resistance to overfitting, and interpretability through feature importance metrics.

Datasets
The datasets frequently used in this domain include clinical features like age, cholesterol levels, chest pain type, blood pressure, and more aligned with your dataset. These features are critical indicators of cardiovascular health and form the basis for most predictive models.

Key Findings
Studies have consistently validated the effectiveness of the a robust ensemble algorithm in predicting heart disease with high accuracy. By utilizing clinical datasets comprising features such as age, chest pain type, cholesterol levels, and others, a robust ensemble algorithm models excel in identifying potential heart conditions.

In this research, the use of a robust ensemble algorithm ensures not only accuracy but also interpretability through feature importance analysis. This aligns with contemporary trends in AI-driven techniques research, which emphasize robust and practical solutions for healthcare applications.

IV. METHODOLOGY

The method employed in this study was to apply the a robust ensemble algorithm algorithm to predict heart disease based on selected clinical parameters. The steps undertaken for this process include the following:

Data Collection and Preparation: The dataset of Heart_Disease_Prediction.csv consists of important features such as Age, Chest pain type, BP, Cholesterol, Max HR, ST depression, Number of vessels Fluro, and Thallium.

The target variable is Heart-Disease. The data has been checked for missing values before considering it for model training to ensure data's integrity and consistency

Feature Selection: Relevant features were selected based on their importances for prediction of heart diseases. Evidence of their importance to heart disease capture important clinical information that translates well to cardiovascular health.

Model Selection: a robust ensemble algorithm was preferred as it is flexible and can handle most types of datasets, it models the interactions of important features well, and it gives good performance with very high interpretability.

Training and Testing: The dataset was split into training and testing sets using 80-20 ratios with the train_test_split method to prevent bias from evaluation.

The model was trained on the training dataset and tested for accuracy on the testing dataset.

Model Evaluation: The performance was evaluated using the metrics of precision, recall, F1-score, and accuracy using a classification report.

The model that was trained could make very good predictions, which make a model ready for clinical decision support.

Model Deployment: The trained model is serialized using Python's pickle library and stored as heart disease prediction. model for future use in real-time applications.

V. EXPERIMENTAL SETUP

The experimental setup for this research includes the following components and processes:

- **Environment:** The implementation was carried out using python, and primarily the following libraries were used:
- **Pandas:** For manipulation and analysis of data.
- **Sklearn:** For implementing and evaluating AI-driven techniques models.
- **Pickling:** For serialization and storage of models.
- **Configuration of Hardware:** The experiments were run on a system with the following specifications:
- **Processor:** Intel Core i5/i7 or its equivalent. RAM of at least 8 GB.
- Required disk space to save the dataset and serialized model.
- **OS:** Windows/Linux/MacOS.
- **Dataset Details:** Dataset given here is Heart_Disease_Prediction.csv consists of: Age, Chest pain type, BP, Cholesterol, Max HR, ST depression, Number of vessels fluro and Thallium.

Target Variable: Heart Disease (for binary classification).
Preprocessing: Check and process missing values in order to have a clean data set. Using domain knowledge and validity to heart disease prediction, selected features were based on that.
Model Training: The data was split into a training and test set at an 80:20 ratio using the train_test_split function from

sklearn. The model uses the robust ensemble algorithm Classifier to train using its ensemble learning abilities which handle interactions among features and allow good prediction accuracy.

Assessment of Performance: The performance of the model under training has been evaluated based on the classification metrics concerning the test dataset. Analysis of results was undertaken with a view to ascertaining the model's ability to effectively predict heart disease.

Model Deployment: The model was trained and serialized for deployment purposes using the pickle library, which will meet its intentions in the future for real-world applications.

VI. CONCLUSION AND FUTURE WORK

The robust ensemble algorithm model developed for heart disease prediction performs very well and will very likely find use within clinical practice.

Model Performance: The robust ensemble algorithm Classifier performs an average of 87 percent accuracy on test data, indicating its capability in separating individuals with or without heart disease from one another. Other performance metrics like precision, recall, and F1- score measure too showed very good results, supporting the model's credibility.

Significance of Findings

Model Reliability: The accuracy of 87% demonstrates high predictive power, thus making the model suitable in assisting health practitioners in the identification of at-risk individuals. The ensemble learning approach by a robust ensemble algorithm renders it robust against overfitting and improves model performance on a wider range of datasets.

Feature Contributions: Key impressive features such as Age, Chest Pain Type, BP, and Cholesterol had high contributions to the prediction by a robust ensemble algorithm itself regarding automatic feature importance and interaction handling that underlay the relevance of the approach to the stated problem.

Comparisons with Existing Methods: The achieved accuracy is in line with or surpasses the results of other AI-driven techniques models as documented in the literature, such as those for Logistic Regression or Decision Trees characterized by average accuracy values.

Future Works

- To widen the sample diversity in the dataset to strengthen the robustness of the model.
- To include other clinical and demographic features for better prediction accuracy.

Compare a robust ensemble algorithm with other AI-driven techniques methods by the performance of the models.

Development of a user application for making the model available in clinical work.

REFERENCES

1. AHA (American Heart Association). (2021). Heart Disease and Stroke Statistics—2021 Update: A Report From the American Heart Association. Circulation, 143(8), e254-e743. https://doi.org/10.1161/CIR.0000000000000950.
2. A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method. https://nature.com/articles/s41598 024-74656-2
3. Prediction of Heart Disease Using Random Forest and Rough Set Based Feature Selection: Breakthroughs in Research and Practice. https://www.researchgate.net/publication/331724547_Prediction_of_Heart_Disease_Using_Random_Forest_and_Rough_Set_Based_Feature_Selection_Breakthroughs_in_Research_and_Practice.
4. Heart Disease Prediction using Random Forest Classifier https://ieeexplore.ieee.org/document/10452459 .
5. Prediction of Heart Diseases using Random Forest https://iopscience.iop.org/article/10.1088/17426596/1817/1/012009/pdf.